DOCUMENT RESUME

ED 057 467 EA 003 862

AUTHOR	Levine, Donald M.
TITLE	Structuring Program Analysis for Educational Research.
PUB DATE	Jul 71
NOTE	20p.; Paper presented to Joint Study Group on Military Research Allocation Methodology (JSGOMRAM) (26th, Burlingame, California, November 17, 1970)
AVAILABLE FROM	Communications Department, Rand, 1700 Main Street, Santa Monica, California 90406 ($1.00)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Cost Effectiveness; Decision Making; Educational Planning; *Educational Research; Evaluation Methods; *Performance Criteria; Program Effectiveness; *Program Evaluation; Program Planning; Systems Analysis
IDENTIFIERS	*Planning Programming Budgeting Systems; PPBS

ABSTRACT

Education is too complex, multidimensional, and poorly understood to lend itself to a single cost/effectiveness criterion. Rather, analysts in the educational field seek to rank alternatives by their effectiveness, report separately on the cost implications, and leave the tradeoffs to the decisionmaker's judgment. In this context, a program is any specifiable mix of components, including pupil and neighborhood characteristics. Program effects are judged by comparison with preprogram status. These effects are evaluated against the objectives, and the results are translated into measures of effectiveness, taking into account interdependencies as necessary and changing the weighting of elements with magnitude. (Author)

ED057467

N-CO
EA

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

STRUCTURING PROGRAM ANALYSIS FOR EDUCATIONAL RESEARCH

Donald M. Levine

July 1971

EA 003 862

P-4565

ED057467

STRUCTURING PROGRAM ANALYSIS FOR EDUCATIONAL RESEARCH

Donald M. Levine*

The Rand Corporation, Santa Monica, California

In most areas of the social sciences, and particularly in education, two fundamental facts obstruct approaches to rational decision-making. First, objectives usually are vague and have not been formulated systematically. No structure of objectives exists within most organizations devoted to social concerns--the objectives of different levels of the organization and of functional units at the same levels are confused with and duplicate each other. Hence, intra-structure communication and coordination can be poor. Moreover, we find that a poor understanding of objectives produces such poor effectiveness criteria that we sometimes do not know what we mean by success. There is hardly any need to expand this point: The problem encountered in assessing and comparing the merits of different educational approaches is quite generally understood. And if we are confused about the meaning of success, we know neither how to evaluate nor how to describe the performance of our system.

The second and less immediate problem is concerned with our understanding of the input-output or causal relationships in the social sciences. Essentially this means that even if we knew what we wanted in terms of educational output--that is, even if our objectives and criteria

---

*Any views expressed in this paper are those of the author. They should not be interpreted as reflecting the views of The Rand Corporation or the official opinion or policy of any of its governmental or private research sponsors. Papers are reproduced by The Rand Corporation as a courtesy to members of its staff. Dr. Levine is a Consultant to The Rand Corporation and Assistant Professor, Ontario Institute for Studies in Education, Toronto, Canada.

This paper was presented to the 26th Joint Study Group on Military Research Allocation Methodology (JSGOMRAM) as part of an all-day panel on "New Directions in Educational Planning," Burlingame, California, November 17, 1970.

Special thanks are due Mr. Richard P. Junghans, Harvard University for his helpful assistance during the preparation of this paper. Throughout, he acted as discussant and critic and was very useful in the development and clarification of several of the topics discussed here.

EDCATA

were clear--we would not know how to design a system to meet our needs. We expect that lowering the student/teacher ratio will improve student achievement, but limited funds are generally available and we have to balance the effect of increasing teachers and thus the number of classrooms in use versus that of hiring reading specialists, for example. Each of these factors must be dealt with in its own way if we are to improve educational planning.

The Planning-Programming-Budgeting System (PPBS) has been developed, in part, to aid in the clarification of this first issue. The bureaucratic structure is organized around the concept of the program. Units are organized around specific objectives that fit in some rational way into a coordinated hierarchical conception of system goals. Financial data are related to the goals and goal achievement of each unit. Reports summarizing and assessing the activities of each unit and each level within the structure are produced at regular intervals. Evaluation of these reports allows for a refinement and reorganization of system activities to better satisfy system needs. Objectives are reexamined and recast; methods of performance measurement and evaluation are subjects of ongoing research.

The implementation of such systems as PPBS is in itself a large topic. We shall focus rather on analysis feasible within the context of such a system.

Within the program structure, objectives are the justification for *an* expenditure of resources; they are the *raison d'etre* of a program. It is around the objectives that a program is developed. It must be understood, however, that objectives are not intrinsic to a program; a program functions quite independently of the objectives. Simply attributing the genesis of a program to a certain set of goals does not bias its function in favor of achieving those objectives any more than if the program was originally established under wholly different auspices.* This functional independence of programs and objectives makes

---

*Note also C. J. Hitch and R. N. McKean, *The Economics of Defense in the Nuclear Age* (New York: Atheneum, 1967), p. 159. "But while good intentions are sometimes reputed to be excellent paving materials, in themselves they do not pave the way to preferred action."

it necessary to utilize measures of effectiveness in judging the relative worths of competing alternative programs.

The program may be viewed as a "black box" that, once organized, performs operations on the input functions, e.g., students as described by factors of intelligence, socio-economic level, race, age, beginning achievement level, etc., such that the output functions generated preserve some of the original factors while changing others. Other influences remaining constant, we attribute these changes to effects of the program. The criteria, in effect, seek out these changes and map them into an effectiveness scheme where the value of the changes is judged against what we set out to accomplish. This judgment is tempered by the cost or resource requirements of the program. By establishing a continuum of such effectiveness-cost positions, we could in theory select the most effective program for a given cost or the least expensive program for a given effectiveness.

An important dichotomy in the use of the term "criterion" arises in the literature. The first is a stand of judging the merit of a program in terms of both its cost and effectiveness to provide a final basis for selection among alternatives--the so-called "test of preferredness." The second is a standard by which the effectiveness of different programs is evaluated and ranking of effectiveness positions is achieved. The former view is supported by Quade* and Hitch and McKean,† whose theoretical treatments are the basis for much of modern practice, but we find that people closer to applications are developing their own definition. It is the latter view that we will adopt.

The principal difficulty of the "selection among alternatives" or "test of preferredness" usage of "criterion" lies in the implicit maximization of net benefits. In cost-benefit analysis it may be possible to achieve this by maximizing the present value difference of the discounted stream of revenue and the discounted stream of costs. Consistent with the theory of the firm, inputs are increased until the shadow price

---

*E. S. Quade (ed.), *Analysis for Military Decisions*, The Rand Corporation, R-387, November 1967.

†Hitch and McKean, op. cit.

equals unity. The characteristic feature of cost-benefit analysis is just this commensurability (after discounting) of input and output measures; all costs and benefits are measured in terms of their dollar value. In such situations, it may be meaningful to stipulate selection of the program yielding the greatest net benefit--this is a criterion according to the strict sense of the word.

To measure net benefits, however, we must have a means of insuring commensurability among all the factors to enter consideration. This is accomplished relatively simply in systems to which we might apply cost-benefit analysis, but such transformations seem inconceivable in the context of most other systems; this is particularly true of problems that we encounter in education, where inputs are often measured in dollars and outputs in behavioral terms. Even if methods were available to achieve commensurability among these items, it would be a mistake to combine the inputs and outputs into a final rating measure since this would tend to obscure the high degree of certainty associated with the cost figures and other input data without enhancing that of the output measures.

In practice, the analyst rarely goes beyond obtaining separate figures for cost and effectiveness of alternatives; the decisionmaker normally supplies his own criterion for making the final selection. The final step of the sort of decisionmaking is difficult to systematize as it usually relies on a highly intuitive grasp that is simultaneously cognizant of the background and intended future of a problem. The decisionmaker is usually content to apply his personal and, hence, largely implicit criterion at this point, and, beyond helping to sharpen the decisionmaker's intuition, the analyst is well advised to avoid this last step. The work of the analyst is complete when the decisionmaker is at ease with the problem.

By presenting two measures--cost and effectiveness--we obviate the necessity of entering the highly qualitative area of the final judgment. We derive another advantage from our definition of criterion: The dichotomy of costs and benefits is explicitly recognized and contrasted within the framework of the program. Not only is it possible to say, "We pay X and we get Y," but our evaluation has produced the

estimate, Y, that is a judgment of how effective the program has been in achieving our stated objectives. And thus cost-effectiveness analysis is introduced (contrast Figs. 1 and 2).

We will consequently understand the criterion to be a "measure of effectiveness," but, as yet, a fully operational definition is not developed. This we will come to shortly.

Effectiveness is a scale of the level of goal attainment; it is the degree to which a program succeeds in fulfilling the objectives. Performance, however, is a scale of the *effects* of a program, and, despite any positive connotations, "performance" in this usage is strictly descriptive. The *performance* of a system is judged against the *objectives*, and evaluation yields the *effectiveness*. Performance

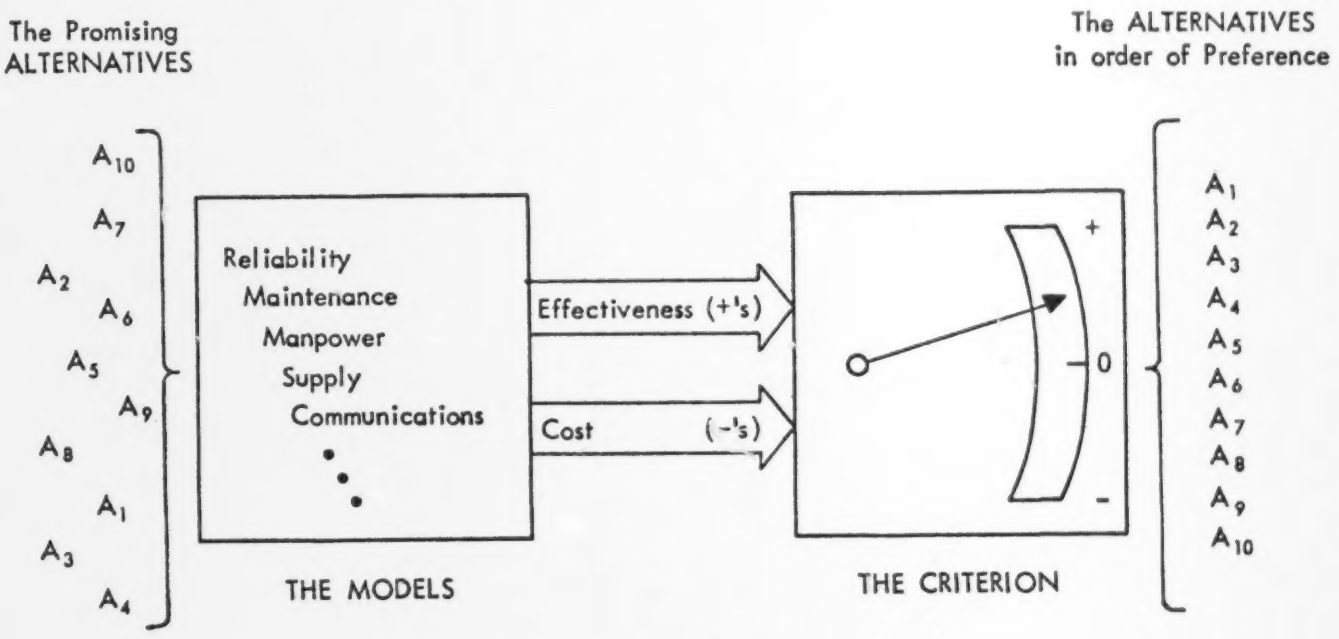


Fig. 1--The structure of analysis, from E. S. Quade, Systems Analysis Techniques for Planning-Programming-Budgeting, *The Rand Corporation, P-3322, March 1966.*

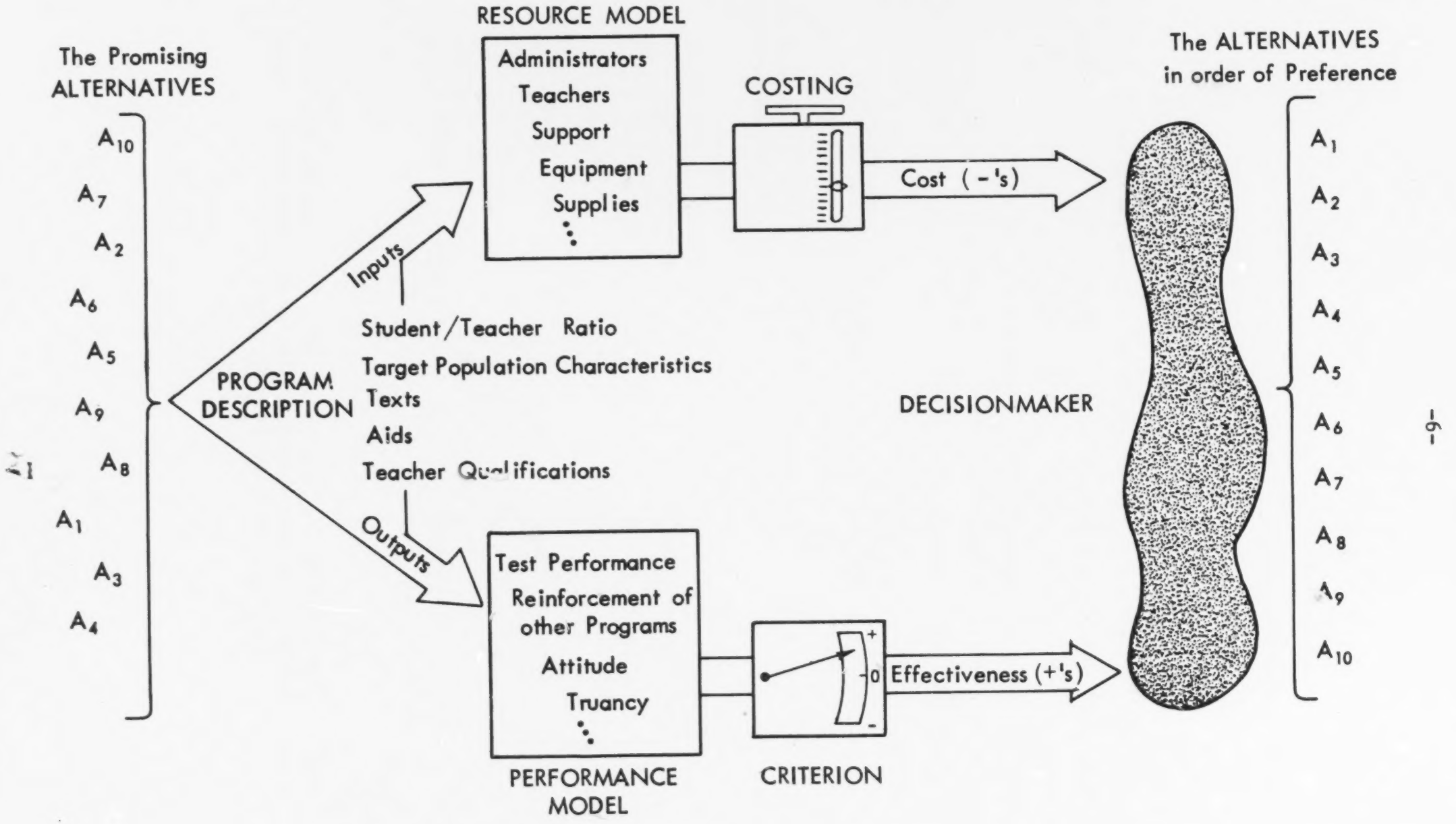


Fig. 2--The structure of cost-effectiveness analysis

is a vector quantity determined by the magnitude assumed by the program on different "effect" axes; effectiveness, on the other hand, is a scalar. Between the assessments of performance and effectiveness, these effect axes are weighted according to their relative importances to arive at a single value that characterizes how well objectives have been met. The unidimensionality of effectiveness now permits us to rank our alternative programs in terms of objective achievement. (Ranking among higher-dimensional quantities is not meaningful without prior scalar assignment.)

The evaluation of performance is motivated primarily by our need to assess the effectiveness of a program in achieving objectives, and in this sense this description is only a means to an end. Inasmuch as performance cannot be measured in and of itself (in that no clearly appropriate units or scales of "performance" exist), we must utilize *measures of performance* to represent in operational or behavioral terms selected effects of a program; within statistical limits, the values of performance measures are assumed to represent the axes of a program's performance description. In turn, these performance measures derive their values from interpretation of tests or *performance indicators*.

Consider a program alternative for the improvement of reading skills among average third-graders. Pre- and post-tests are administered and, by analysis of the differences, measures of system performance are gained. In the same way that we "test" the system for performance, we will "test" performance for effectiveness. The plane of evaluation has shifted. No longer are we examining the program with a view to describing its effects; our present concern is to judge the value of these effects against the objectives that we originally sought to attain by installation of the program. This is the role of the criterion: It is a "test of effectiveness" (as opposed to "preferredness"), a rule by which we measure the effectiveness of a program in terms of its performance characteristics.

The criterion is an interpretation of the objectives that assigns the weights that each performance measure (performance is typically a function of several variables) should receive in contributing to the overall goal achievement. But the criterion is more than a simple

combination rule to be followed in assigning effectiveness to a set of performance measures; it is usually variable in addition. Invoking the so-called "law of diminishing returns," as a desired item becomes increasingly available, each additional increment in its availability becomes less desirable. In some cases, even monotonicity cannot be assured. Our criterion must account for such nonlinearities when they assume significant dimensions. In summary, then, the criterion is an expression of the objectives that:

1. Recognizes
   a. the relative importance of different performance characters in contributing to the overall effectiveness of a program, and
   b. how that weighting changes as a function of the magnitude assumed by each performance measure (the projection of the performance vector on each of the "effect" axes).
2. Utilizes this knowledge in transforming the performance of a system into a final valuation of effectiveness.*

---

*By translating into mathematical terms, the conceptual relationship between performance, criteria, and effectiveness is quite simply demonstrated. Performance is an n-dimensional vector, $\overline{P}$, which is composed by performance measures, p_1, p_2, p_3, ..., p_n. The criterion is a $1xn$ matrix, $\overline{\overline{C}}$, with coefficients, c_1, c_2, c_3, ..., c_n. The coefficients, c_i, are weighting factors for the values assumed by the performance measures, p_i, in producing contributions to the overall effectiveness of the program, E(more properly, $E\overline{I}$, where $\overline{I}$ is a one-dimensional identity vector). That is,

$$\overline{\overline{C}}\ \overline{P} = E\ \overline{I} = E,$$

where $\overline{\overline{C}}$ = Criterion matrix,
$\overline{P}$ = Performance vector,
$\overline{I}$ = Identity vector (= 1),
E = Effectiveness scalar,

$$(c_1, c_2, c_3, \ldots,)\begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ \vdots \end{pmatrix} = c_1p_1 + c_2p_2 + c_3p_3 + \ldots, = E\ \overline{I} = E.$$

Let us consider a simple example. An examination consisting of three parts--Vocabulary, Speed Comprehension, and Grammar--is administered to assess how "effectively" a person reads. This is to be the basis for evaluating different vocational reading courses whose stated purpose is to improve reading at an everyday, practical level. Our task is to utilize these scores in some way to arrive at an effectiveness valuation by which the alternative programs can be ranked. We will assume equal initial performance levels for each of the groups.

Each of these parts would be a performance indicator in the context of this example. The actual test scores would be the values *assumed* by the performance indicators, which, in turn, would be translated into performance levels to be assumed by each of the performance measures. In mathematical terms, the values finally associated with the performance measures would correspond to the projection of the reading performance vector onto the separate "effect" axes of Vocabulary, Speed Comprehension, and Grammar, as shown in Fig. 3.

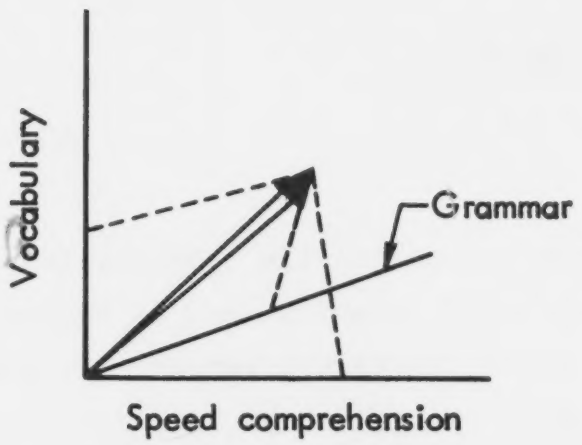


Fig. 3--Projection of reading performance vector

In itself, of course, a test score possesses no intrinsic value: It is in the context of the examination that the score takes on meaning. It is the interpretation of these scores, then, that yields the performance. For Vocabulary, the performance has a relatively simple meaning--the number of words in a person's vocabulary--but for Speed Comprehension, and particularly Grammar, a more subjective description is required. For simplicity, we will suppose for the moment that numerical values can be reasonably associated with different "levels"

of performance with respect to each of the "effect" axes or performance measures (see Table 1).

Table 1

NUMERICAL VALUES OF PERFORMANCE
RELATED TO TEST SCORE

Performance Measure	*Test Score*	*Performance*
Vocabulary	97-100	15000 + (words)
	90-96	14000 +
	89-89	12500 +
	60-77	10000 +
	41-59	9000 +
	below 40	below 9000
Speed Comprehension	89-100	10
	81-88	9
	67-80	8
	56-66	7
	38-55	6
	below 38	5
Grammar	93-100	10
	84-92	9
	76-83	8
	53-75	7
	30-52	6
	below 30	5

The criterion recognizes the relative importance of the different factors in the activity, reading, and it becomes the rule by which the scale positions of the different performance measures are combined into a final valuation of effectiveness. If all the programs have the same budgetary requirements, the choice could be based on the effectiveness ranking. If cost factors differ, then the decisionmaker must balance the cost and effectiveness valuations of each of the programs to arrive at a final selection (see Table 2).

While our example may have served the purpose of illustrating the relationships between some elements of systems theory in education, it is at the same time very crude. By inclusion of relevant statistical data on the performance indicators, we might have been able to bound more precisely the performance levels for each performance measure. The criterion itself would have to have been the result of considerable

Table 2

CONVERSION OF PERFORMANCE MEASURES TO FINAL VALUATION OF EFFECTIVENESS

Course Code	*Performance Indicator*	*Score*	*Performance Measure*	*Performance Level*	*Effectiveness*
A	Vocab	80	Vocab	12500 (words)	11.2
	Speed Comp	50	Speed Comp	6	
	Grammar	53	Grammar	7	
B	Vocab	60	Vocab	10000 (words)	13.4
	Speed Comp	81	Speed Comp	9	
	Grammar	64	Grammar	7	
C	Vocab	79	Vocab	12500 (words)	12.2
	Speed Comp	64	Speed Comp	7	
	Grammar	66	Grammar	7	
D	Vocab	69	Vocab	10000 (words	12.6
	Speed Comp	79	Speed Comp	8	
	Grammar	78	Grammar	8	

Ranking of courses: B, D, C, A, where the coefficients of the criterion matrix are

$$c_1 = 0.0003$$
$$c_2 = 1.0$$
$$c_3 = 0.2.$$

study. First, the performance levels would have been somehow standardized to indicate what "normal" means and, second, the relative importance of each of the performance measures would have been stated within the context of the need to read at an everyday, practical level. The coefficients of the criterion itself might be statistically described. In this way, we could test the significance of our conclusions ranking the effectiveness of the different programs.

By assuming that the coefficients of the criterion were constant, we neglected the possibility of interdependencies. In education we are particularly plagued with the complexity of the relationships influencing performance. Very often such interdependencies must be explicitly recognized if our models are to be of much predictive value.

Independence essentially means that variation with respect to one performance measure does not affect the contribution that another makes

in the assessment of effectiveness. If performance measures are interdependent, then we are not only required to (a) assign weights that each performance measure is to receive in achieving the effectiveness assignment, but also to (b) determine how this weighting scheme should adjust itself for different levels of achievement with respect to each of the performance measures. As the interdependency decreases, the need to perform the latter analysis vanishes.

This makes sense. If A and B, as before, are performance measures and if B satisfies a need 40 percent as effectively as A, then surely exceptionally high performance with respect to B would decrease the urgency of achievement with respect to A. Installation of playground facilities for physical education, for example, may serve as sufficient attitudinal stimulus for pupils that investment in the form of new teachers to decrease the size of classes might reasonably be delayed. Presumably, the state of the student-teacher ratio would become correspondingly less important, and this would be reflected in the criterion for effectiveness encompassing these programs. This is need interdependency. If A and B are need independent, then performance on B does not affect the weight that A should receive; the needs they directly fulfill are disjunct. This type of interdependency was included in our definition of the criterion on page 8.

A second type of interdependency is important in a different context. If the dependencies are more direct, that is, if altering system A affects the performance of system B, investigation of the interactions can be more complex. This is functional interdependency. We might consider the effect of a school's health program on its overall performance in interscholastic athletics. In many instances, this type of interdependency is not of much consequence: When comparing distinct programs, whatever the interdependency of their components, it is the outputs we examine for evaluation of effectiveness, and the particular means by which performance was stimulated is not of immediate concern. In planning, however, this type of interdependency is important since the variables we encounter are continuous or many-valued, and data are largely matter for speculation; we need some basis for predicting the

consequences of different policies in program design.*

Returning to the example of the previous section, we see that we have to determine need interdependencies and that the more complex functional interdependencies--those that relate the contributions of grammar to speed comprehension and vocabulary--are not essential in making a selection of alternative programs. Of course, our task is somewhat reduced, but there still exists the considerable problem of associating different levels of performance on each of the performance measures with the higher objective "to assess reading ability."

The discussion of analysis so far has outlined the approach to selection among a finite number of alternative programs on the basis of effectiveness. Measures of output quantities are available or readily

---

*The two kinds of independence are easily illustrated. Recall that the criterion is the weighting scheme for the transformation of performance (measures) into effectiveness, where $c_1p_1 + c_2p_2 + \ldots c_np_n$ is the effectiveness. There are essentially two ways of altering the sum of the c_ip_i: Change the c_i (as in need interdependence) or change the c_i (as in functional interdependence).

(a) Need interdependence. If systems A and B are need interdependent, this is reflected in the criterion. The weighting factor for the first system, c_1, would be a function of p_2 as well as p_1. Similarly, c_2 would be a function of p_1. If A and B are need independent, then neither is c_1 a function of p_2 nor is c_2 a function of p_1.

(b) Functional interdependence. Functional interdependence of systems A and B implies that p_1 is a function of p_2 and *vice versa*; that is, $p_1(p_2, \ldots)$ and $p_2(p_1, \ldots)$. Functional interdependence does not affect the criterion.

It is clear that dependencies in (a) will always be interdependent; that is, if c_1 is a function of p_2, c_2 will be a function of p_1. It is not so apparent, however, that the same will apply in (b). It might be said that one system can function quite satisfactorily independent of the state of the first, while the first system is dependent on the performance of the second. The logic is deceptive; to cite a level of performance of system A may determine that of system B, but, conversely, citing a performance level for system B constrains system A to perform to a particular level. It may seem more appropriate to label one variable "controlling" and another "dependent," but our point of reference is in fact arbitrary.

predictable. But what of the need to design programs in the first place? How are we to anticipate the effectiveness of a particular program? How do we discover "efficient" solutions to our problems--that is, the most effective solution at a fixed cost or the least expensive solution to attain a given effectiveness?

The criterion yields a measure of effectiveness, but it relates the outputs of a program to objective attainment and does not tell us any information about the input control variables that is required for planning purposes. Consequently, it now becomes necessary to understand the relationship of inputs (in terms of operational characteristics rather than cost requirements) to outputs and the role of functional interdependencies as introduced earlier. These interrelations must be stated explicitly to delimit the feasible solutions at specific cost levels, and consequently modelling becomes important in the planning context of cost-effectiveness analysis.

Through the performance model, we here relate our decision variables--the student-teacher ratio, for example--to the results of performance indicators, which in turn are related to the effectiveness through performance evaluation. In this way, we are able to evaluate the effectiveness of different policy decisions. Strictly speaking, then, while it is necessary to construct cost and effectiveness curves for the different performance measures, they are *output* quantities and such graphs are only of passing interest. During the ensuing discussion, input decision variables will be the performance characteristics.

We shall embark on a brief discussion of effectiveness curve analysis more from the standpoint of illustrating the gulf that exists between current planning theory and practice than as a system subject to ready application. Rigorous development of these techniques requires a more exact knowledge of the criteria and interdependencies than is presently available in education.

Conceptually, we formulate the objective function, f (the criterion for effectiveness), and the cost function, g, in terms of the performance characteristics, x_i. A new function, h, is formed such that $h = f + \lambda g$. The Lagrange function, h, is differentiated with respect to each of the x_i and λ, and the partial differential equations are each

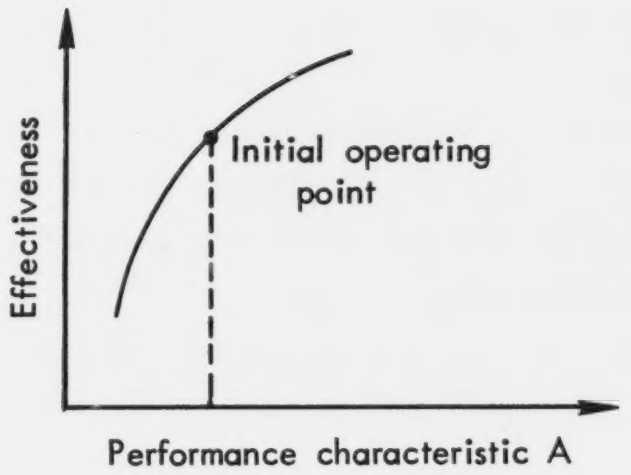
set to zero. The group of equations are solved simultaneously, eliminating λ, and the result yields the set of variable values that correspond to the optimal system performance for the given cost level. We are no longer evaluating a small number of alternative programs to satisfy a particular set of objectives; instead, we are reviewing a potentially infinite number of combinations of policy alternatives.

An equivalent and slightly more qualitative statement of the same process involves the development of five graphs (see Fig. 4) for each performance characteristic:

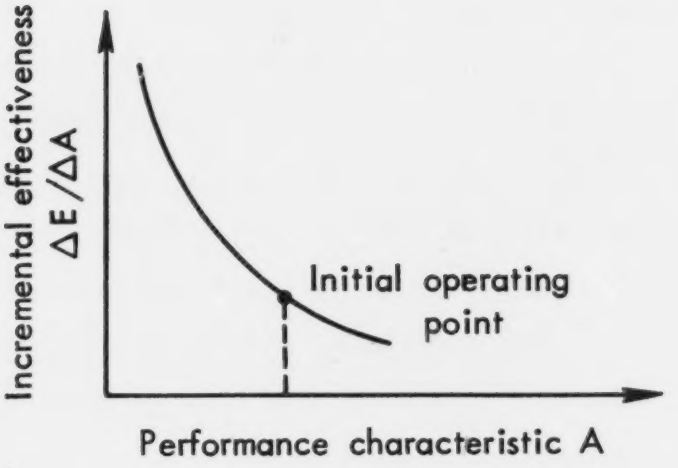
1. Effectiveness versus performance.
2. Cost versus performance.
3. Incremental effectiveness versus performance.
4. Incremental cost versus performance.
5. Incremental effectiveness/incremental cost versus performance.

Graph 5 is a plot of Lagrangian multiplier values for different system characteristic performance positions. It is derived from graphs 3 and 4 which, in turn, follow from graphs 1 and 2. Variables are perturbed until equal marginal gain positions are attained, and an efficiency curve is developed to guide policy decisions. For the chosen effectiveness level, the decisionmaker knows that the system is at lowest cost and, conversely, at peak effectiveness for a fixed cost level. At this stage he must decide which operating level best suits his needs.

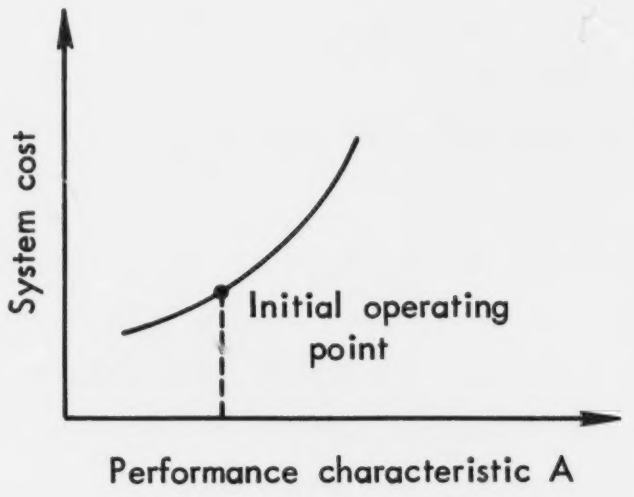
Unfortunately, education is a long way from using even this more qualitative Lagrangian approach to planning. In general, the input-output processes in education are too poorly understood to be modelled adequately; a great many variables need to be standardized before causal relationships can fairly be deduced. The performance model and the criterion are independent, but if the output data supplied by the model (to assign values to the performance measures) are incorrect, the criterion will supply a correspondingly inaccurate view of the effectiveness of the hypothetical program. However, the lack of good models need not deter us from the application of systematized decisionmaking in education. The primary role of theoretical models is to supply us



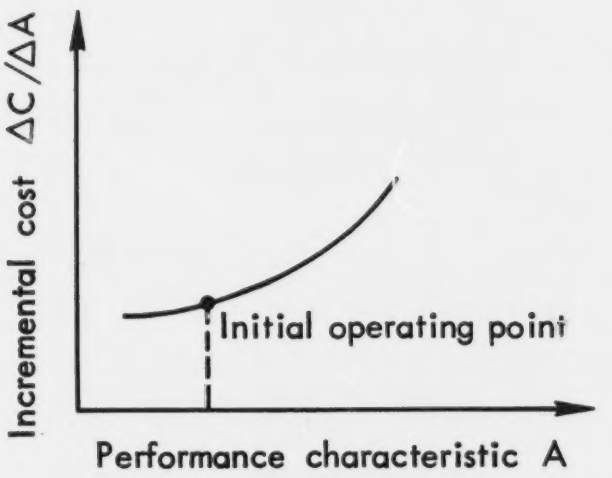
Graph 1



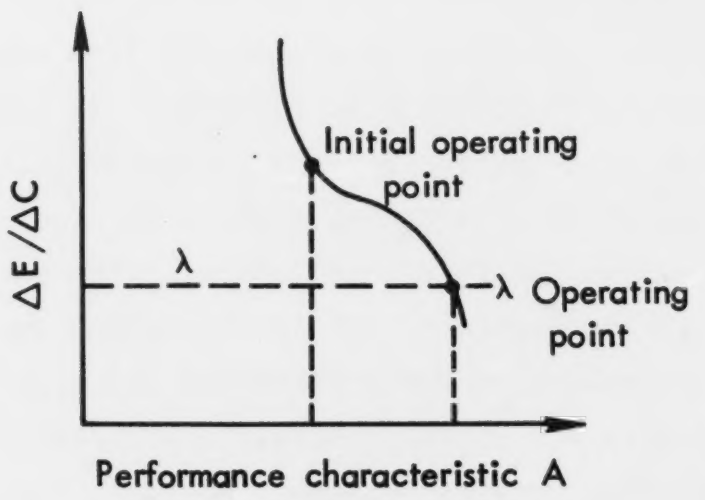
Graph 3



Graph 2



Graph 4



Graph 5

Fig. 4--The derivation of cost-effectiveness parameters

with data about nonexistent programs; if we could just select the best from existing programs, we should have made great progress.

A program is any specifiable mix of system components that controls outputs in a way relevant to the assessment of effectiveness. Classes of 20 and 21 students, although not greatly different with respect to size, are different programs. The same teacher, the same instruction materials and techniques, the same physical plant, the same student-teacher ratio, the same distribution of IQs among students--this program in an upper-class neighborhood and this program in a ghetto are different programs; the effects of similar educational environments on pupils from different socio-economic backgrounds are not likely to be the same. We find the concept of the program very central to much of the analysis that we would wish to undertake. The program defines an activity unit for which the relationship between data elements is assumed significant. Depending on the purpose at hand, we define in different ways the bounds of what we would call a program.

Techniques of operations research would guide us to better planning were we only able to model the educational process itself. While effectiveness curve analysis may not presently be available, nothing prevents us from evaluating discrete programs in education: No performance models are required and functional interdependencies can be neglected. This provides a great reduction of the problem while significant gain is still to be anticipated. In this country there are 19,000 public school districts containing a total of 120,000 elementary and secondary schools with 50,000,000 pupils and 2,000,000 teachers. Thousands of different programs (as we have defined the term) are currently in existence presenting numerous examples for nearly every conceivable mix of system characteristics at levels well above what is required for statistical significance. We might ask, What is the best program to institute in a ghetto school for eighth-grade science? Using the methods of performance and criteria evaluation and collecting data on existing programs, we evaluate the effectiveness and cost of the "alternatives." With these data, the administrator can make the best selection commensurate with his budgetary limitations. The ultimate refinement of the system would provide programs tailored to the

particular student population and teacher characteristics for each subject of study, but, initially, we might just provide program suggestions adapted to the needs of the average ghetto school science teacher.

In a planning context, we would like to examine questions of organization and timing of curricula. We would like to assess the relative merit of instructional aids and teaching methods. We would like to investigate noncurricular quantities: the effect of the student-teacher ratio on the achievement of different age groups, the effect of different physical education programs on student morale in the academic areas, the effect of various guidance programs on the truancy rate. If we could answer these questions, we could improve education, but education itself--its modes and objectives--is not well understood.

By formalizing the data and analysis structures, however, we can work towards planning models while utilizing the power of this system to aid in selection among alternative existing programs; analysis can be a research tool as well as an evaluative technique. Because of the large-scale and the scope of the effort implied, the framework of program budgeting or program analysis is appropriate for the investigation of the broader educational principles on which we might base a gradua revamping of the entire educational system. The data required are essentially those needed for a systematic evaluation of many alternative programs for a particular educational niche, except that we are interested in many such niches.

The information needed is of two types: (1) descriptive input data to permit grouping of data sources by policy types and (2) output data to serve as performance indicators for the evaluation of effectiveness. Data are grouped into classifications by target population characteristics, curriculum composition, pupil-teacher ratio, teacher qualifications, and so on, and the analysis of standardized test scores and other performance indicators is subsequently performed. A multidimensional (performance) data matrix is constructed--the cells of the matrix labeled by input characters and occupied by output measures--and analysis is performed for different combinations of cells to reduce successively the influence of different input factors. In this way, we can focus on individual factors over different ranges of the remaining

input variables and derive *in quantitative terms* the input-output relations in education and the functional interdependencies of the system components. Performance models with statistical bounds are now established for each of the subsystem components.

If, however, our first attempt to derive these models should yield too many relationships that are not statistically significant or that are too complex to analyze in a short time, we still have a powerful decision aid at our disposal. The criterion of a given social group is applied to the cells of the matrix in the regions relevant to its own situation to achieve a multidimensional effectiveness matrix. A resource model is developed to provide accompanying cost data. In this way, the decisionmaker gains a view of the potential of existing alternative programs (with costs) to meet his needs. When satisfactory models *are* achieved, we can use them in conjunction with the effectiveness curve analysis to improve educational systems (and resource allocations) even further.

These educational relationships are sufficiently complex that casual policy analysis is not a satisfactory approach to resolving the different factors. Specific issues, such as whether to invest funds in halving the student-teacher ratio or to develop new physical facilities, could be subjects for policy analysis, but the more fundamental questions require investigation on a grander scale. School systems must be geared for intensive introspection and analysis if the confusion in educational planning is to be allayed; the Planning-Programming-Budgeting System provides an appropriate framework in which to approach both basic research and policy analysis.